The Episodic Master Recession (EMR) Program for R

Lara Mitchell, John Nimmo, Charlie Horowitz

March 27, 2014

This article documents code written to implement the Episodic Master Recession (EMR) method (Nimmo et al. 2014) in the R environment (R Core Team 2013). The EMR method identifies discrete episodes of recharge at a hydrological site based on the water-table dynamics and precipitation record at the site during a given time period. Episodes are identified by rapid increases in the water-table level, and a variation on the Water Table Fluctuation method (Heppner and Nimmo 2005) is used to estimate recharge on an episodic basis. Each episode is paired with a single precipitation event likely to have contributed to the water-table rise.

The EMR program consists of two functions, `applyEMR` and `plot.EMR`. The function applyEMR carries out routine calculations and data formatting procedures, identifies recharge episodes and their correlated precipitation events, calculates recharge, and outputs an “EMR” object summarizing the results. The function `plot.EMR` takes an “EMR” object as input and generates plots to help the user visualize the results. These functions were built under R version 3.0.2.

Method Overview

The EMR method requires four time series:

1. time, T (as a numerical quantity, e.g., in days)
2. water-table level, H
3. cumulative precipitation, P
4. water-table rate-of-change, dH/dT

The method depends on two parameters that characterize the site: (1) the water-table fluctuation tolerance (δ_T), essentially a measurement-noise criterion used to ascertain whether or not a given fluctuation in the water-table level is hydrologically significant, and (2) the precipitation lag time (t_l), which reflects the response time for recharge caused by a given precipitation event.

Generally, an episode is defined as a period during which the observed water-table rate-of-change, dH/dT, exceeds the Master Recession Curve (MRC)-predicted water-table rate-of-change, dH/dT_{MRC}, by an amount greater than the fluctuation tolerance. Visually, this means that the dH/dT vs. T curve is above the fluctuation tolerance curve, defined by (dH/dT_{MRC} + fluctuation tolerance) vs. T. The actual start time for the episode is defined to be the time at which dH/dT last crosses dH/dT_{MRC} prior to crossing above the tolerance curve OR one lag time before the tolerance crossing, whichever occurs closer in time to the tolerance crossing. The end time for the episode is defined to be the time at which dH/dT first crosses dH/dT_{MRC} after crossing back under the tolerance curve OR one lag time after the tolerance curve re-entry, whichever occurs closer in time to the return crossing. Since all of these curves are defined by a finite number of data points, curve-crossing times are estimated by linearly interpolating between the points immediately before and after the crossing. If one episode starts before the previous episode ends, the episodes are combined into one episode. The precipitation event associated with an episode begins one lag time before the start of the episode, and ends one lag time after the end of the episode.

Potential episodes can be discarded for the following reasons:

1. There are insufficient data to define a start time and/or an end time. This can happen if, for example, dH/dT crosses above dH/dT_{MRC} but does not re-enter during the time series.

2. There are insufficient data to define the corresponding precipitation event, i.e., the first potential episode starts 'too early' in the time series.

3. The net precipitation during the precipitation event is less than `precip_bound`.

The recharge associated with an episode is calculated by the following procedure: First, the MRC is used to predict how the water-table level would have behaved in the absence of the precipitation event. This involves extrapolating the water-table level forward in time iteratively, beginning with the water-table level at the start of the episode. The extapolation is carried out until the time $t_{end} - t_l$, where t_{end} is the time at which the episode ends. A similar extrapolation is carried out backward in time from the end of the episode to the time $t_{end} - t_l$. The difference between these two curves at $t_{end} - t_l$, ΔH, is then multiplied by the specific yield to give the episodice recharge: R = SY x ΔH.

`applyEMR` *Apply the Episodic Master Recession (EMR) method.*

Description

This function applies the Episodic Master Recession (EMR) method to a set of hydrological data. and returns a list containing summary data on recharge episodes identified by the method. The output list has the class "EMR".

Usage

```
applyEMR(data_frame, fluc_tol, lag_time, SY, mrc_type, mrc_coef,
         N=0, precip_bound=0, save_file=NULL, delta_t=1,
         include_extrap=TRUE)
```

Arguments

`data_frame`	a data frame containing time, water-table, and cumulative precipitation data. Optionally, it can also contain water-table rate-of-change and precipitation rate-of-change data.
`fluc_tol`	a number giving the water-table fluctuation tolerance. Should be greater than or equal to 0.
`lag_time`	a number giving the precipitation lag time. Should be greater than 0.
`SY`	a number giving the specific yield. Should be greater than or equal to 0.
`mrc_type`	a character string indicating the mathematical form of the Master Recession Curve (MRC). The two options are "polynomial" and "power".
`mrc_coef`	a numeric vector (of length at least 1) giving the coefficients of the MRC. For polynomials, the coefficients should be in increasing exponent order. See 'Details'.

`N`	a positive integer used as a moving-average smoothing parameter. If N is 0, no smoothing is applied. Defaults to 0.
`precip_bound`	a number giving the minimum amount of precipitation required to define an episode. Should be greater than or equal to 0. Defaults to 0.
`save_file`	a character string giving the name of an optional csv file containing a summary of the function results. The name should include the file extension (.csv). If `NULL`, no file is created. Defaults to `NULL`.
`delta_t`	a number used to control the step size when extrapolating the water-table curves during an episode. Should be greater than 0. Defaults to 1. See 'Details'.
`include_extrap`	logical; if `TRUE`, include water-table extrapolation data in the output (primarily for plotting purposes). Defaults to `TRUE`. See 'Details'.

Details

`data_frame` should be a data frame with either three or five columns (column names are arbitrary). In either case the first three columns must contain time data, T, water-table level data, H, and cumulative precipitation data, P, respectively. All the values of T should be unique, otherwise the function will terminate with an error message. P should be monotonically increasing over time. If a decrease is detected, the function will fill in the appropiate elements with the last precipitation value before the decrease, and give a warning message.

If `data_frame` contains three columns, the water-table derivative with respect to time, dH/dT, is calculated inside the function by:

$$\left.\frac{dH}{dT}\right|_i = \frac{k_1^2 * H_{i+1} + (k_2^2 - k_1^2) * H_i - k_2^2 * H_{i-1}}{k_1 * k_2 * (k_1 + k_2)},$$

where $k_1 = T_i - T_{i-1}$, $k_2 = T_{i+1} - T_i$, and i is an index indicating the i^{th} record. Similarly for the precipitation derivative with respect to time, i.e., the instantaneous precipitation rate, dP/dT. The user has the option of providing dH/dT and dP/dT as the fourth and fifth columns in `data_frame`, respectively. If the user wants to provide dH/dT but not dP/dT, the fifth column should contain all NA's. Similarly, if the user wants to provide dP/dT but not dH/dT, the fourth column should contain all NA's. If dH/dT is provided but has any missing values, dH/dT is recalculated by the program and a warning message is displayed in the console. Simiarly if dP/dT is provided but has missing values. We note that dP/dT is not strictly required for the EMR method, but it is used for informational purposes in the function output. (See 'Values'.)

For each of the two values of `mrc_type`, the table below shows the corresponding mathematical form of the MRC, and the assumed syntax for `mrc_coef`.

`mrc_type`	MRC Mathematical Form	`mrc_coef`
"polynomial"	$\frac{dH}{dT} = p_0 + p_1 H + p_2 H^2 + ...p_n H^n$	$\texttt{c}(\texttt{p}_0, \texttt{p}_1, \texttt{p}_2, \ldots, \texttt{p}_\texttt{n})$
"power"	$\frac{dH}{dT} = p_1 + p_2 (H - p_3)^{p_4}$	$\texttt{c}(\texttt{p}_1, \texttt{p}_2, \texttt{p}_3, \texttt{p}_4)$

An optional moving-average smoothing process can be applied to dH/dT by setting `N` to an integer greater than 0. The j^{th} element of the smoothed water-table rate vector is given by

$$\frac{d\hat{H}}{dT}_j = \frac{1}{N^2} w * h_j^T,$$

where $w = (1, 2, \ldots, N, \ldots, 2, 1)$, $h_j = (H'_{j-N+1}, \ldots, H'_{j-1}, H'_j, H'_{j+1}, \ldots, H'_{j+N-1})$, and H' is used to denote the non-smoothed dH/dT vector. If smoothing is applied, the beginning and end of T, H, P, and dP/dT are trimmed so that all vectors remain the same length.

The temporal step sizes used in the extrapolation processes are equal to the step sizes inherent to T (e.g., $T_i - T_{i-1}$), multiplied by the quantity `delta_t`. Hence, `delta_t` can be used to achieve a finer step size. If `include_extrap` = `TRUE`, the extrapolated time and water-table vectors for each episode are included as part of the function output - mainly to help make plotting easier; see the description of `plot.EMR`. It is optional because the ouput already contains a lot of information and the user may or may not want to keep these data as well.

Minimal argument-checking is carried out at the beginning of the call to `applyEMR`. For example, if any of the arguments `data_frame`, `fluc_tol`, `lag_time`, `SY`, `mrc_type`, and `mrc_coef` are missing, the function is terminated with an error message.

A note on units. No assumptions are made about the units of time, water-table, or precipitation, except that they are consistent across all fields. For example, if T, H, and P are given in days, meters, and millimeters, respectively, then dH/dT (if provided) should be in meters/days and dP/dT (when provided) should be in millimeters/day. The units of `fluc_tol` and `lag_time` are assumed to be the same as the units of H and T, respectively. The units of the specific yield constant, `SY`, should be consistent with the units of H.

Value

A list with class "EMR" and components

`final_data`	a data frame containing the formatted time, water-table, cumulative precipitation, water-table rate-of-change, and precipitation rate-of-change time series. Examples of formatting include "fixing" decreases in precipitation and smoothing the water-table rate (See 'Details'). If any episodes have been identified, these time series are updated to contain linearly-interpolated values corresponding to the episode start and end times, and the precipitation event start and end times.
`parameters`	a named vector containing `fluc_tol`, `lag_time`, `SY`, and `precip_bound`.
`MRC`	a named list containing `mrc_type`, `mrc_coef`, and `mrc_fcn`, where `mrc_fcn` is the MRC function.
`episode_data`	a data frame containing a summary of each episode, including episode start and end times, precipitation event start and end times, estimated recharge values, and net episodic precipitation. If there are no episodes, then `NULL`.
`fwd_extrap`	if `include_extrap` is `TRUE` and there are episodes, a list with each element containing the forward extrapolation times and water-table levels for a single episode. Otherwise `NULL`.
`bwd_extrap`	if `include_extrap` is `TRUE` and there are episodes, a list with each element containing the backward extrapolation times and water-table levels for a single episode. Otherwise `NULL`.

If no episodes are identified, `episode_data`, `fwd_extrap`, and `bwd_extrap` are `NULL`. If `include_extrap` = `FALSE`, `fwd_extrap` and `bwd_extrap` are `NULL` regardless of the number of episodes.

A complete listing of the fields in `episode_data` is shown below. Note that `avg_precip_rate` and `max_precip_rate` are extracted from the precipitation rate-of-change time series in `final_data`.

`episode_num`: episode number
`start_time`: episode start time
`end_time`: episode end time
`duration`: time duration of the episode
`recharge`: estimate of episodic recharge
`start_precip_time`: precipitation event start time
`end_precip_time`: precipitation event end time
`start_precip`: cumulative precipitation value at `start_precip_time`
`end_precip`: cumulative precipitation value at `end_precip_time`
`net_precip`: net precipitation during the episode

`avg_precip_rate`: average precipitation rate during precipitation event
`max_precip_rate`: maximum precipitation rate during precipitation event
`fwd_extrap_start_time`: time at which forward water-table extrapolation begins
`fwd_extrap_end_time`: time at which forward water-table extrapolation ends
`fwd_extrap_start_H`: water-table value at the start of forward extrapolation
`fwd_extrap_end_H`: water-table value at the end of foward extrapolation
`bwd_extrap_start_time`: time at which backward water-table extrapolation begins
`bwd_extrap_end_time`: time at which backward water-table extrapolation ends
`bwd_extrap_start_H`: water-table value at the start of backward extrapolation
`bwd_extrap_end_H`: water-table value at the end of backward extrapolation

Examples

...

plot.EMR *Generate plots using an EMR object.*

Usage

```
plot.EMR(x, plot_type="mrc", t_label="Time", H_label="H",
         dH_label="dH/dt", p_label="Cumulative precip",
         b1="lightsteelblue2", b2="steelblue"
         leg_x="topleft", leg_y=NULL, cex.lab=1.3, cex.axis=1.2, ...)
```

Arguments

`x`	an EMR object.
`plot_type`	a character string giving the type of plot to be generated. `"mrc"` produces a plot of dH/dT vs. T. `"hydro"` produces a plot of H vs. T. Defaults to `"mrc"`.
`t_label`	a character string giving the desired time label (x-axis). Defaults to `"Time"`.
`H_label`	a character string giving the desired water-table label (y-axis), for `plot_type="mrc"`. Defaults to `"H"`.
`dH_label`	a character string giving the desired water-table rate label (y-axis), for `plot_type="hydro"`. Defaults to `"dH/dt"`.
`p_label`	a character string giving the desired precipitation label (right-hand y-axis). Defaults to `"Cumulative precip"`.
`b1`	a character string giving the color to be used for cumulative precipitation. Defaults to `"lightsteelblue"`.
`b2`	a character string giving the color to be used for precipitation events. Defaults to `"steelblue"`.
`leg_y`	a number giving the y-coordinate of the legend OR `NULL`.
`cex.lab`	a number giving the size of the axis labels (same as in `plot()`).
`cex.axis`	a number giving the size of the axis values (same as in `plot()`).
`...`	further arguments and graphical parameters.

Details

`"mrc"` figures also show the fluctuation tolerance band, MRC, cumulative precipitation, recharge period(s), and precipitation event(s). `"hydro"` figures also show the cumulative precipitation, recharge period(s), precipitation event(s), and episode extrapolations (if available). This function can still be used when there are no episodes.

Examples

...

Author(s)

Lara Mitchell

References

Heppner, C.S. and Nimmo, J.R. 2005. A computer program for predicting recharge with a master recession curve. U.S. Geological Survey Scientific Investigations Report 2005-5172.

Nimmo, J.R., Horowitz, C., and Mitchell, L. 2014 Discrete-storm water-table fluctuation method to estimate episodic recharge. *Groundwater* DOI: 10.1111/gwat.12177.

R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, 2013. URL http://www.R-project.org/.